AD________________

Award Number: W81XWH-11-1-0744

TITLE: Development of Assays for Detecting Significant Prostate Cancer Based on Molecular Alterations Associated with Cancer in Non-Neoplastic Prostate Tissue

PRINCIPAL INVESTIGATOR: Farhad Kosari, Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
ROCHESTER, MN 55909

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. REPORT DATE** October 2014

**2. REPORT TYPE** Annual

**3. DATES COVERED** 30 Sep 2013 - 29 Sep 2014

**4. TITLE AND SUBTITLE**
Development of Assays for Detecting Significant Prostate Cancer Based on Molecular Alterations Associated with Cancer in Non-Neoplastic Prostate Tissue

**5a. CONTRACT NUMBER**

5b. GRANT NUMBER
W81XWH-11-1-0744

5c. PROGRAM ELEMENT NUMBER

**5d. PROJECT NUMBER**

**6. AUTHOR(S)**

Farhad Kosari F, Cheville JC, Vasmatzis G, Karnes RJ, Manemann M, Murphy SJ

E-Mail: kosari.farhad@mayo.edu

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Mayo Clinic
Rochester, MN 55905

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT** The goal of this project is to develop biopsy based assays to assess the probability that patients with a negative biopsy or with a prostate cancer (CaP) Gleason score 6 (GS6) biopsy actually have "significant" CaP of Gleason score 7 or higher which was missed during the biopsy evaluations due to insufficient sampling. In 2014, we considerably expanded the size of the discovery samples which will improve the robustness of biomarkers and developed methodologies for improving the detection of epigenetic modification in rare cells. Also, we tested several methods for multiplex analyses of transcriptome and DNA methylation changes. Further, we have secured additional funding which will allow us to expand transcriptome and epigenome profile of indolent and significant prostate tumors. Finally, we have applied for major grant initiatives which if funded, will help expanding the scope of this study through the use of nano-scale devices and quantum dot based assays. We are expecting the publications of our research results in the coming year.

**15. SUBJECT TERMS**
LCM AACR

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** USAMRMC |
|---|---|---|---|---|---|
| **a. REPORT** U | **b. ABSTRACT** U | **c. THIS PAGE** U | UU | 11 | **19b. TELEPHONE NUMBER** *(include area code)* |

**Table of Contents**

**Introduction**:

Each year in the U.S. more than a million men with an elevated serum PSA or abnormal digital rectal exam undergo a prostate biopsy, and nearly 200,000 are found to have prostate cancer (CaP). Decisions to treat CaP are heavily influenced by the Gleason score (GS) of the tumor in the needle biopsy specimen. Gleason score is a measure of tumor differentiation based on the two most prevalent patterns of tumor growth. Patients whose entire tumor is composed of GS6 rarely progress, and recently, more men diagnosed with GS6 tumors on needle biopsy are selecting active surveillance rather than surgery or radiation therapy. In contrast, men with more poorly differentiated tumors (GS7 and higher) have a significantly increased risk of progression, and require treatment. Choosing the best treatment options for patients with biopsy GS6 is complicated by the fact that a biopsy procedure only samples a very small part of the prostate, and in about 30% of men, it underestimates the GS [1]. In those cases, men with GS7 and higher ($GS7^+$) prostate cancer are assumed to have GS6 tumors potentially leading to inappropriate treatment. In addition, because of the limited sampling and 30% false negative rate for detecting cancer [2,3], many men with a negative biopsy result may have clinically significant prostate cancer. Because of that, many of the 800,000 patients with a negative biopsy undergo repeat biopsies which can be frustrating for both patients and urologists. When a pathologist examines a prostate needle biopsy specimen, the focus is on the identification of prostate cancer and appropriate Gleason scoring. Very little attention is paid to the "normal" areas which often comprise the majority of biopsy samples. This is despite a considerable body of evidence suggesting that molecular alterations associated with tumor in adjacent non-neoplastic cells, the so called "tumor field effect", can provide valuable clues regarding the status of the tumor. Remarkably, the field effect (FE) alterations have also been associated with aggressive prostate cancer [4].

**Body / Results:**

The objective is to develop clinically relevant molecular models to predict significant prostate cancer with $GS7^+$ based on the prostate cancer FE markers. This proposal will focus only on identification of significant tumors with $GS7^+$ because Gleason score is the single strongest predictor of outcome in men with prostate cancer, and has the greatest influence on the clinical management of men with prostate cancer. This proposal will concentrate on the "omics" areas where prostate cancer FE has been best demonstrated, namely transcriptomic and epigenomics. There are two Aims. Aim I will identify and validate prostate cancer FE markers associated with $GS7^+$ tumors. Aim II will develop and test molecular models for predicting upgrading in GS6 biopsies and for predicting $GS7^+$ cancer in a repeat biopsy.

Table I describes four types of samples for which genomic profiles are to be generated in this research and the initial size of the samples. As described below, the sample size in each category has been expanded. The 4 types of samples included non-cancerous tissues from CaP patients with (i) indolent GS6 CaP ($N_{i6}$ = 5), (ii) GS3+4 CaP ($N_{3+4}$ = 5), and (iii) GS 8 and higher CaP ($N_{8+}$ = 5). We also analyze benign

**Table I:** Bulk and LCM samples proposed in the application for the biomarker discovery step by the next generation sequencing

| Sample | Bulk | LCM (HGPIN) |
|---|---|---|
| BP | 5 | |
| $N_{i6}$ | 5 | 5 |
| $N_{3+4}$ | 5 | 5 |
| $N_{8+}$ | 5 | 5 |
| **Total** | **20** | **15** |

prostate tissues from patients free of CaP (BP = 5) as controls. BP samples are resected prostate tissues from patients who were not diagnosed with CaP but had their prostates resected in cystoprostatectomy operations because of bladder cancers. In the proposal for the first phase of the project, gene expression and epigenetic alterations are to be analyzed by next generation sequencing. Laser captured microdissection (LCM) is used to collect high grade prostatic intraepithelial neoplasia (HGPIN) lesions in 15 samples. The remaining samples are collected using bulk macro-dissection (Table I).

**Research Accomplishments**:

Collection of BP tissues and selection of cases for the next generation sequencing (NGS): In April 2014, we requested and received approval for expanding the discovery set from the original plan and for no cost extension. Table 2 describes the status of the collected samples that have been sequenced or are in the queue for library preparation and sequencing. It also describes the samples for which this DOD funding is used and the target sample numbers. We were able to secure additional funds from the Prostate SPORE and Center of Individualized Medicine (CIM) Clinomics Program at the Mayo Clinic to expand the samples without incurring additional cost to this DOD proposal. Additionally, we collaborated with Dr. Thibodeau who also holds a DOD grant for transcriptome sequencing of bulk BP samples. It is important to note that genomic analyses of HGPIN and benign bulk samples funded by this DOD grant will not be delayed by the collection of data in tumor and associated normal samples.

**Table 2:** Bulk and LCM samples collected in this study for genomic profiling. Numbers in parenthesis indicate number of samples for which funds from this DOD proposal were used or will be used for generating the sequencing data. Numbers in the brackets are the targets for the project

| | **Bulk** | **LCM** | | |
|---|---|---|---|---|
| **Sample** | **benign Areas** | **HGPIN** | **aN** | **tumor** |
| **BP** | 19 (0) [19] | | | |
| **$N_{i6}$** | 20 (8) [20] | 13 (20) [20] | 10 (0) [10] | 13 (0) [20] |
| **$N_{3+4}$** | 10 (7) [10] | 9 (10) [10] | 9 (0) [10] | 16 (0) [30] |
| **$N_{8+}$** | 20 (12) [20] | 19 (20) [20] | 10 (0) [10] | 13 (0) [20] |
| **Total** | 69 (27) [69] | 41 (50) [50] | 29 (0) [30] | 42 (0) [70] |

In coordination with the Tissue Request Acquisition Group (TRAG) at the Mayo Clinic, resected prostates from cystoprostatectomy patients (BP) for this project have been collected.

Collection of samples and DNA/RNA isolation: Cases with adequately large areas of HGPIN for LCM collections and benign areas that did not include any preneoplastic or tumor regions by macro-dissection in each of the $N_{i6}$,

$N_{3+4}$, and $N_{8+}$ categories were identified in our frozen prostate tissue registry. Additionally, we required that sufficient quantities of tumor be available for separate LCM collection. All samples in this study were collected under IRB approved protocols.

Under the supervision of Dr. Cheville, LCM samples have been collected by a meticulous and carefully implemented process to minimize degradation of nucleic acids, especially the RNA. Purified RNA from LCM samples by a Qiagen kit was of high quality (RIN numbers typically greater than 8) and quantity (tens of nano-grams). Also DNA from LCM samples by an Arturus or a Qiagen kit was of sufficient quantity for the RRBS library prep. DNA and RNA were simultaneously collected from the bulk samples by the AllPrep kit (Qiagen) and had very high quality and yield.

Library preparations for transcriptome profiling: We have been working closely with the the Mayo Genomics Facility (MGF) in preparing the libraries for RNA-seq. All LCM RNA-Seq libraries have been using 5 nanograms starting material and producing uniform output by NuGen protocol. Similarly, all bulk samples produce high quality TruSeq libraries.

Library preparations for methylation profiling: In 2013, we worked out a protocol for methylation profiling by reduced representation bisulfite sequencing (RRBS) using the HGPIN samples collected by LCM. To increase the sensitivity of the RRBS for detecting epigenetic changes by the CaP FE in a fraction of cells in the bulk samples, in a collaborative effort and with funding support from the MGF we have been testing a modified RRBS protocol. The original RRBS protocol [5] uses methylation insensitive MspI to digest genomic DNA and is an effective technique to identify moderate and large methylation differences between samples. For detecting differences that are subtle and of small magnitude as we expect to see in the study of cancer field effect, however, RRBS has inherent limitations. These limitations largely originate from the fact that methylated and unmethylated genomic DNA fragments are both present and amplified together in the same library. However, in our project, it will be important to separate these two libraries and thereby increase the signal to noise levels. Also, these two types of templates have drastic differences in the ‘GC’ content and amplify with different efficiencies. Amplifying them together results in selective enrichment of un-methylated DNA, a phenomena which has long been recognized as ‘PCR bias’ [6].

We proposed a simple modification to the RRBS protocol. The idea was based on sequential digestion of DNA first with GlaI which only cuts methylated DNA and then with MspI which cuts the remaining and largely unmethylated DNA. In this way, methylated and unmethylated DNA libraries are processed separately and can be sequenced with different indexes at coverages that are dependent on the project priority. A similar strategy was proposed by by Jelinek [7]. However, Jelink protocol does not offer single CpG resolution that is offered by our protocol. An RRBS library by GlaI (Figure 1) produced largely methylated

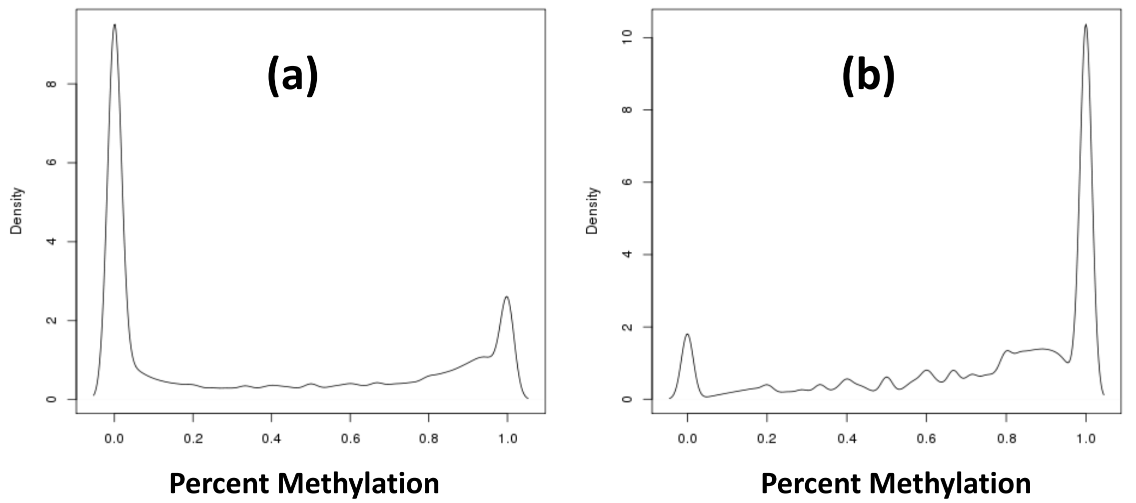



**Figure 1**: Percent methylation of the RRBS libraries generated by MspI (as in standard protocol) (a), and by GlaI (b). GlaI only cuts methylated genomic DNA and thereby sequence fragments in (b) are enriched for the percent methylation.

DNA and had 92.3% mappable fragments. This was compatible with standard RRBS library by MspI which had 89.4% mappable fragments. We plan to use this strategy for generating RRBS libraries from the bulk samples.

<u>Processing of the transcriptome data</u>: We have been developing a bioinformatics algorithm that focuses in identification of recurrent splice variants that distinguish $N_{8+}$ from $N_{i6}$. These biomarkers can be incorporated in sensitive PCR assays where even small fractions of cells expressing the splice variants can be readily detected. To increase the possibility of success, we sequenced bulk samples at high depth (3 samples per lane) and long (100 bp) sequence fragment lengths. A new hire, Nikolas Vasmatzis, has been dedicating his time to this project and his efforts have generated promising results.

For example, a recurring exon 4-7 splice junction in EYA4 gene (Figure 2) was detected in 92% of $N_{8+}$ and only in 25% of $N_{i6}$.(Table 3). Table 3 describes splice junctions with best discrimination between $N_{8+}$ and $N_{i6}$. This list includes prominent cancer related genes, such as TP73, ERBB2, and HDAC10. We are in the process of validating these results in independent samples.

**Figure 2**: A newly developed bioinformatics program identified recurring 4-7 splice junction in EYA4 in 92% of $N_8$. By contrast, only 25% of $N_{i6}$ samples had this junction (Table 3)

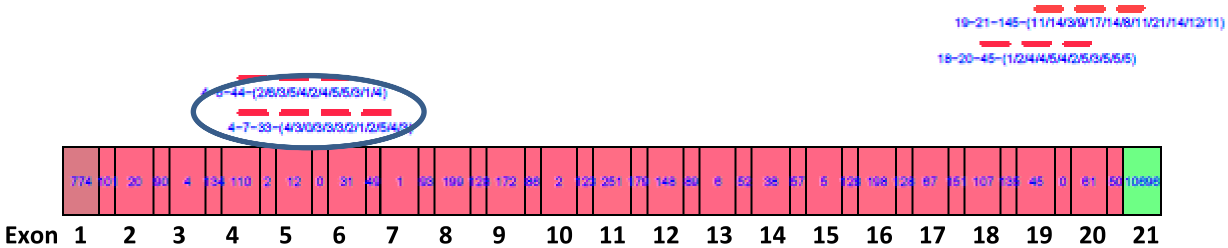


**Table 3** Recurrent splice junctions in $N_{8+}$ (left panel) and $N_{i6}$ (right panel)

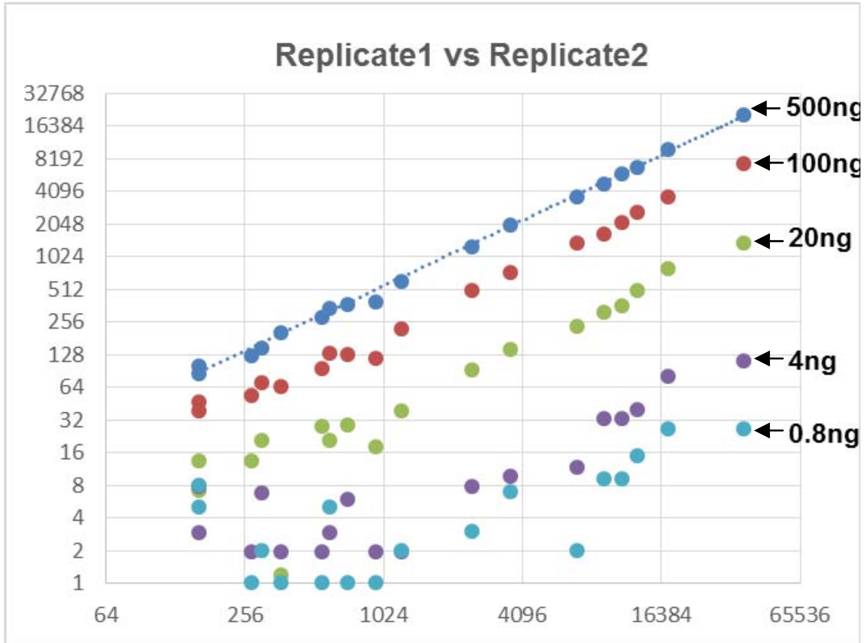
| Recurrent in $N_{8+}$ | | | | | | Recurrent in $N_{i6}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Gene_Name** | **Junction** | **N8** | **Ni6** | **Delta** | | **Gene_Name** | **Junction** | **N8** | **Ni6** | **Delta** |
| EFEMP1 | 5_7 | 75.0 | 0.0 | 75.0 | | NT5DC2 | 1_3 | 25.0 | 100.0 | -75.0 |
| EYA4 | 4_7 | 91.7 | 25.0 | 66.7 | | SLC44A2 | 13_16 | 8.3 | 75.0 | -66.7 |
| MKLN1 | 13_15 | 66.7 | 0.0 | 66.7 | | RQCD1 | 7_10 | 0.0 | 62.5 | -62.5 |
| SEC61A2 | 12_14 | 66.7 | 0.0 | 66.7 | | TMTC1 | 10_12 | 25.0 | 87.5 | -62.5 |
| SEMA6C | 17_19 | 58.3 | 0.0 | 58.3 | | CTCF | 2_5 | 0.0 | 62.5 | -62.5 |
| NIM1 | 1_3 | 83.3 | 25.0 | 58.3 | | CWC25 | 2_4 | 0.0 | 62.5 | -62.5 |
| SYK | 2_4 | 83.3 | 25.0 | 58.3 | | PRKACA | 1_3 | 25.0 | 87.5 | -62.5 |
| ADK | 7_9 | 83.3 | 25.0 | 58.3 | | FEZ2 | 7_9 | 16.7 | 75.0 | -58.3 |
| TBX5 | 7_10 | 58.3 | 0.0 | 58.3 | | CCM2 | 6_8 | 41.7 | 100.0 | -58.3 |
| C1orf151-NBL1 | 1_3 | 66.7 | 12.5 | 54.2 | | ANK3 | 31_33 | 16.7 | 75.0 | -58.3 |
| FAM13A | 21_23 | 66.7 | 12.5 | 54.2 | | M6PR | 1_4 | 16.7 | 75.0 | -58.3 |
| ITGA7 | 21_23 | 91.7 | 37.5 | 54.2 | | GSTZ1 | 8_10 | 41.7 | 100.0 | -58.3 |
| ABCD4 | 17_19 | 91.7 | 37.5 | 54.2 | | **HDAC10** | 12_14 | 8.3 | 62.5 | -54.2 |
| **TP73** | 3_6 | 75.0 | 25.0 | 50.0 | | AGTR1 | 3_5 | 33.3 | 87.5 | -54.2 |
| TMEM51 | 1_4 | 75.0 | 25.0 | 50.0 | | DIAPH1 | 8_10 | 8.3 | 62.5 | -54.2 |
| SEMA6C | 8_10 | 75.0 | 25.0 | 50.0 | | SLC37A3 | 2_5 | 33.3 | 87.5 | -54.2 |
| CP | 10_12 | 50.0 | 0.0 | 50.0 | | NCOA4 | 11_14 | 8.3 | 62.5 | -54.2 |
| USP4 | 13_15 | 75.0 | 25.0 | 50.0 | | METTL1 | 3_5 | 8.3 | 62.5 | -54.2 |
| HERC3 | 13_15 | 75.0 | 25.0 | 50.0 | | **ERBB2** | 16_18 | 8.3 | 62.5 | -54.2 |
| PDLIM4 | 1_4 | 75.0 | 25.0 | 50.0 | | EML2 | 12_14 | 8.3 | 62.5 | -54.2 |

Another transcriptome gene selection strategy focuses in differentially expressed genes between the two groups. Once all RNA-Seq data become available, we will be able to move forward with the selected genes to the validation experiments.

<u>Methodology development in FFPE surgical samples for validation experiments</u>: As in our previous report, all case selections including biopsies for Aims 1 and 2 are completed. Several strategies for multiplex analyses of gene expression in FFPE samples were tested, including one-step and two-step RT-PCR, gene specific RT-PCR, and NanoString. At the end, the best performer was NanoString. With this technology, it was possible to produce results that were consistent in 500, 100, and 20 ng starting material (Figure 3). This sensitivity makes multiplexed gene-expression analyses in biopsies feasible as we typically obtained 100 ± 60 ng RNA (mean ± SD) by processing of the 2-3 sections of biopsy cores.

<u>Methylation specific PCR</u>: A methylation specific PCR was developed and tested in FFPE samples for detection of GSTP1 methylation in prostate tumors [8]. The assay was remarkably sensitive and capable of detecting methylation changes in tumor compared with normal samples in starting material compatible with prostate biopsies (5-10ng in our experience).

**Figure 3:** Nanostring experiments with varying starting DNA from FFPE blocks. Each point is the expression level of a gene (total of 18 genes). Blue, red, and green dots represent starting DNA of 500, 100, and 20 ng, respectively. Acceptable correlation was observed with starting DNA of tens (> 20ng) to hundreds of nano-grams




**Additional related research activities:** In addition to the steps described above, we have engaged in other research activities which will enhance our abilities to accomplish PC100553 goals. These included:

Secured funding for genomic analyses of indolent and significant CaP: Based in our previous experience [9], FE biomarkers can be robust if biomarker selection strategy is guided by changes in the corresponding tumor. In 2014, we secured a $50,000 award for generating transcriptome and epigenetic profiling of the iG6 and $G7^+$ tumors. This funding will also help in developing a model to predict upgrading of GS6 biopsies and complement models based on FE markers in the benign tissue.

Addition of two new members to the group: Nafiseh Janaki, M.D. is a talented new research fellow who has interest in a pathology residency program. She has been instrumental in organizing the discovery and the validation experiments, and especially assisting Simone Terra, M.D. in the collection of appropriate cells by the LCM. Nikolas Vasmatzis is a bioinformatics programmer with a bachelor degree in computer science and has been developing *de novo* algorithms for mapping transcriptome data and for detecting recurring splice junctions in $N_{i6}$ and $N_{8+}$. He will also assist in selecting the transcriptome and epigenetic biomarkers.

Pending external applications: In a collaborative effort with engineers in the University of Illinois at Urbana-Champaign (UIUC), we have applied for two major grants (R33 and U54). The R33 and part of U54 aim to develop models for stratification of indolent from significant CaP at the biopsy stage. In these applications we plan to develop sensitive nano-scale IR resonators to identify different cell types in a prostate tissue, including reactive stroma and myofibroblast which are modified in the CaP by the field effect. Additionally, selected biomarkers will be tested in a multiplexed fashion by quantum dots.

Un-anticipated events: In 2014, Dr. Rochelle Arvizo was away for a period of time due to the maternity leave. Also, our laboratory moved to a different building in the fall 2014. These events injected some un-anticipated delays in our plans.

**Reportable Outcomes**

In 2014, we considerably expanded the size of the discovery samples which will improve the robustness of biomarkers. We are close in completing the collection of samples in the expanded discovery set and move on with completing genomic profiling and validating field effect biomarkers. We developed methodologies for improving the detection of epigenetic modification in rare cells. Also, we tested several methods for multiplex analyses of transcriptome and DNA methylation changes. Further, we have secured additional funding which will allow us to expand transcriptome and epigenome profile of indolent and significant prostate tumors. Tumor biomarkers will help in improving the accuracy of FE models in diagnosis of a missed significant prostate cancer at the biopsy stage. Finally, we have applied for major grant initiatives which if funded, will help expanding the scope of this study through the use of nano-scale devices and quantum dot based assays.

**Conclusions:**

In 2015, we expect to see the results of this research. The size of the genomic profiling in the discovery step has been expanded which will improve the robustness of our biomarker panels. In the coming year, we will be working on the following publications: (1) correlates of significant and indolent prostate cancers in HGPIN by integrated genomic analyses (2) detection of epigenetic alterations in rare cells by a modified RRBS protocol, (3) a probabilistic molecular model for diagnosis of a missed significant cancer in benign biopsies. These will build on our previous publications in the field [9,10]. Additionally, through further collaborations and team science with scientists and engineers in the Mayo Clinic and other institutions, we hope to bring this exciting research into clinic and improve the care of patients.

**References:**


1. Griffin CR, Yu X, Loeb S, et al. Pathological features after radical prostatectomy in potential candidates for active monitoring. J Urol 2007;178:860-3; discussion 3.
2. Patel AR, Jones JS, Rabets J, DeOreo G, Zippe CD. Parasagittal biopsies add minimal information in repeat saturation prostate biopsy. Urology 2004;63:87-9.
3. Stewart CS, Leibovich BC, Weaver AL, Lieber MM. Prostate cancer diagnosis using a saturation needle biopsy technique after previous negative sextant biopsies. J Urol 2001;166:86-91; discussion -2.
4. Malins DC, Gilman NK, Green VM, et al. Metastatic cancer DNA phenotype identified in normal tissues surrounding metastasizing prostate carcinomas. Proc Natl Acad Sci U S A 2004;101:11428-31.
5. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 2011;6:468-81.
6. Moskalev EA, Zavgorodnij MG, Majorova SP, et al. Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. Nucleic Acids Res 2011;39:e77.
7. Jelinek J, Liang S, Lu Y, et al. Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. Epigenetics 2012;7:1368-78.
8. Baden J, Adams S, Astacio T, et al. Predicting prostate biopsy result in men with prostate specific antigen 2.0 to 10.0 ng/ml using an investigational prostate cancer methylation assay. J Urol 2011;186:2101-6.
9. Kosari F, Cheville JC, Ida CM, et al. Shared gene expression alterations in prostate cancer and histologically benign prostate from patients with prostate cancer. Am J Pathol 2012;181:34-42.
10. Murphy SJ, Cheville JC, Zarei S, et al. Mate pair sequencing of whole-genome-amplified DNA following laser capture microdissection of prostate cancer. DNA Res 2012;19:395-406.